High-dimensional derivative-free optimization via trust region surrogates in linear subspaces

Damien van de Berga, Nilay Shaha, Antonio del Rio-Chanonaa,\*

aSargent Centre for Process Systems Engineering, Roderic Hill Building South Kensington Campus. London, SW7 2AZ, United Kingdom

a.del-rio-chanona@imperial.ac.uk

Abstract

Maintaining the benefits of derivative-free optimization in higher-dimensional decision spaces presents challenges for existing optimization methods. We introduce CUATRO\_PLS - an extension of the CUATRO quadratic trust region optimizer that leverages intrinsic structures across high-dimensional black-box variables. CUATRO\_PLS shows competitive convergence with leading derivative-free optimization algorithms in three high-dimensional chemical engineering case studies even in the absence or underestimation of known intrinsic dimensionality and is significantly faster than other model-based derivative-free optimization algorithms.

**Keywords**: black-box optimization, reduced-order optimization, dimensionality reduction

* 1. Introduction

Derivative-free optimization (DFO) is an effective tool for optimizing chemical engineering systems without explicit gradient expressions. However, state-of-the-art (SOTA) DFO solvers usually scale poorly in the number of decision variables and often even in their evaluation budget (van de Berg et al., 2022). Many existing methods for high-dimensional black-box optimization, from linear-quadratic trust region (Cartis and Roberts, 2023) to Bayesian optimization (BO) (Wang et al., 2016; Letham et al., 2020) methods, rely on random (Gaussian) subspace projections to find embeddings over which to construct and optimize surrogates. While end-to-end learning of linear (Garnett et al., 2014) or nonlinear embeddings (Moriconi et al., 2019) with Gaussian processes or variational autoencoders is possible, limited evaluations can compromise performance.

In this work, we introduce CUATRO\_PLS, which integrates end-to-end learning concepts into the scalable Convex qUAdratic Trust Region Optimizer (CUATRO). At each iteration, Partial Least Squares regression (PLS) identifies the linear embedding that best predicts the output of all evaluations within a trust region. Surrogate fitting and minimization are then performed within this subspace and the minimization candidate is evaluated after reconstruction in the original space. This approach maintains the tractability of working with convex quadratic surrogates in low dimensions while enhancing the quality of the subspaces as compared to Gaussian projections by incorporating response information.

We benchmark CUATRO\_PLS against a random baseline (Latin Hypercube Sampling) and seven SOTA DFO solvers: Py-BOBYQA, DIRECT-L, HEBO, ALEBO, TuRBO, and CMA-ES. We first use a synthetic toy problem with up to 1000 dimensions and a true linear latent space to show that CUATRO\_PLS shows consistent convergence to a high-accuracy solution and is multiple orders of magnitude faster than other model-based methods. We then assess all algorithms on three high-dimensional chemical engineering problems exhibiting intrinsic structure: reactor control policy optimization (61 dimensions), value chain coordination under privacy considerations (72 dimensions), and bi-level planning-scheduling optimization (172 dimensions).

In Section 2, we discuss latent embeddings in DFO and introduce our algorithm CUATRO\_PLS, before describing our benchmarking problems in Section 3. In Section 4, we then show that CUATRO\_PLS ranks among the top three algorithms on all problems and outperforms competitors in high-dimensional scenarios favoring exploitation over exploration.

* 1. Methodology
     1. Derivative-free optimization and latent structure

First, we consider derivative-free optimization (DFO) problems of the following form:

|  |  |
| --- | --- |
|  | (1) |

where denotes the variables within box bounds used to optimize the expensive black-box objective . Underlying our work is the assumption that there is an embedding in the latent space with effective dimensionality capable of explaining or approximating the response of

|  |  |
| --- | --- |
|  | (2) |

where , and is the reconstruction of the latent embedding into the original space such that . The existence of a latent embedding allows us to construct and optimize surrogates in the lower-dimensional space of the black-box objective, “breaking” the curse of dimensionality.

* + 1. High-dimensional derivative-free optimizations

Before introducing our approach CUATRO\_PLS, we present the DFO algorithms we benchmark against.

*Py-BOBQYQA*is a trust-region optimizer constructing linear-quadratic surrogates using interpolation and regression techniques. The algorithm exhibits strong exploitative properties by making quick progress, sometimes at the expense of exploration.

*DIRECT-L* is a search space partitioning direct DFO solver that elegantly trades off exploitation and exploration. The method is more robust to ill-conditioning in the objective as it does not rely on surrogates.

*Latin hypercube sampling* (LHS) is a space-filling design method. Any proposed high-dimensional DFO should outperform random search in the form of LHS.

*CMA-ES*, as an evolutionary search method, shines in highly nonlinear, stochastic applications with a high evaluation budget.

*HEBO*, short for Heteroscedastic and Evolutionary Bayesian Optimization solver, is the winning submission to the *NeurIPS 2020 Black-Box Optimisation challenge*.

*TuRBO* is a high-dimensional Bayesian optimization solver that leverages Thompson sampling within trust regions for scalability.

*ALEBO* revisits some design considerations of previous high-dimensional BO solvers relying on random embeddings.

A considerable part of the high-dimensional Bayesian optimization and DFO literature is predicated on the idea of finding projections to and reconstructions from latent spaces over which to optimize. These methods differ considerably however in the projection and reconstruction techniques used, the choice of surrogates in the latent space and how these are updated. Cartis and Roberts (2023) for example employ random Gaussian projection matrices within trust-region DFO solvers, which improves their scalability on nonlinear least squares problems. While random embeddings preserve theoretical guarantees useful for convergence proofs, we believe that we can integrate response information to improve the sample efficiency of high-dimensional trust region methods.

* + 1. CUATRO\_PLS

We propose CUATRO\_PLS based on CUATRO (van de Berg et al., 2022). CUATRO\_PLS iterates over the following steps until running out of evaluation budget:

1. We sampleinput-output evaluations within the current quadratic trust region of radius r around the trust region center in the original space , until we have evaluations .
2. We then train PLS on all samples within to find the projection and (imperfect) reconstruction matrices that identify the linear embedding from which to best linearly predict such that .
3. We then train convex quadratic surrogates in the lower-dimensional embedding:

|  |  |  |
| --- | --- | --- |
| |  |  | | --- | --- | |  | (3) | |

where is the positive semi-definite matrix defining the quadratic,  
 the linear, and the scalar coefficients.

1. Minimization is then performed within this subspace such that the reconstruction remains in the original trust region :

|  |  |  |
| --- | --- | --- |
| |  |  | | --- | --- | |  | (4) | |

Since we implement our algorithm within a convex programming framework, we relax (4), and define a heuristic trust region in the latent space as follows:

|  |  |
| --- | --- |
|  | (5) |

where = is the furthest distance in the embedding space between the TR center and all samples within.

1. We then evaluate the suggested minimization candidate of (5) after reconstruction to the original space, and go to step 1) if the termination criteria are not met.

For a given latent dimensionality, if we were to use linear, instead of quadratic surrogates in the latent space, PLS would give the optimal embedding. In practice, we find that PLS works well even in combination with quadratic surrogates. The intuition behind this is that the black-box function should display increasingly linear behaviour as the trust region decreases. Next, we describe the case studies used for benchmarking.

* 1. Case studies
     1. High-dimensional Rosenbrock with low effective dimensionality

We first define a synthetic high-dimensional Rosenbrock function with a known linear embedding dimensionality similar to (Wang et al., 2016) with the difference that we mix up the informative and uninformative inputs using an orthogonal Gaussian matrix to make all dimensions informative. Thisallows us to control the original andeffective dimensionality in studying the effect of over- or underestimating the embedding dimensionality used as input to our solver.

* + 1. Chemical engineering applications with low effective dimensionality

We benchmark all algorithms on three high-dimensional chemical engineering problems. While we believe these to exhibit intrinsic structure, the dimensionality of the *true* latent embedding is unknown, and does not have to be linear.  
The *reactor control policy* problem consists in finding the 61 neural network weights and biases that optimize a control policy of a standard CSTR control problem. The parameters of neural networks are known to exhibit low effective dimensionality.

In the *value chain coordination* (van de Berg et al., 2023a), we want to find the material flows (72) that maximize the constituent agents’ planning profits which can only be obtained as a result of privacy-preserving proprietary simulations. We believe that these embeddings are caused by temporal and spatial correlations between material flows.

In the *hierarchical planning-scheduling*, we solve for 172 planning targets that optimize a bi-level planning-scheduling objective as described in (van de Berg et al., 2023b). Intrinsic subspaces are again believed to arise from similar temporal correlations.

* 1. Results and discussion
     1. Synthetic low-dimensional case study

On the 100-dimensional Rosenbrock instance on the top of Figure 1, CMA-ES, ALEBO, and TURBO display the best convergence. CUATRO\_PLS seems to be competitive with other SOTA solvers, with a median and worst-case final convergence comparable to that of CMA-ES. As expected, CUATRO\_PLS has a runtime at least an order of magnitude lower than that of other model-based DFO solvers while displaying comparable performance. HEBO and ALEBO are not investigated further given their exploding runtime on high-dimensional, high-budget problems.

When the dimensionality and budget are then increased to 1,000 on the bottom of Figure 1, we see that CUATRO\_PLS, despite displaying a slightly worse median run than TURBO, displays the best final run of all samples and seems to be the most consistent. Additionally, while TuRBO’s runtime increased by an order of magnitude when doubling the budget, CUATRO\_PLS’s runtime only doubled.

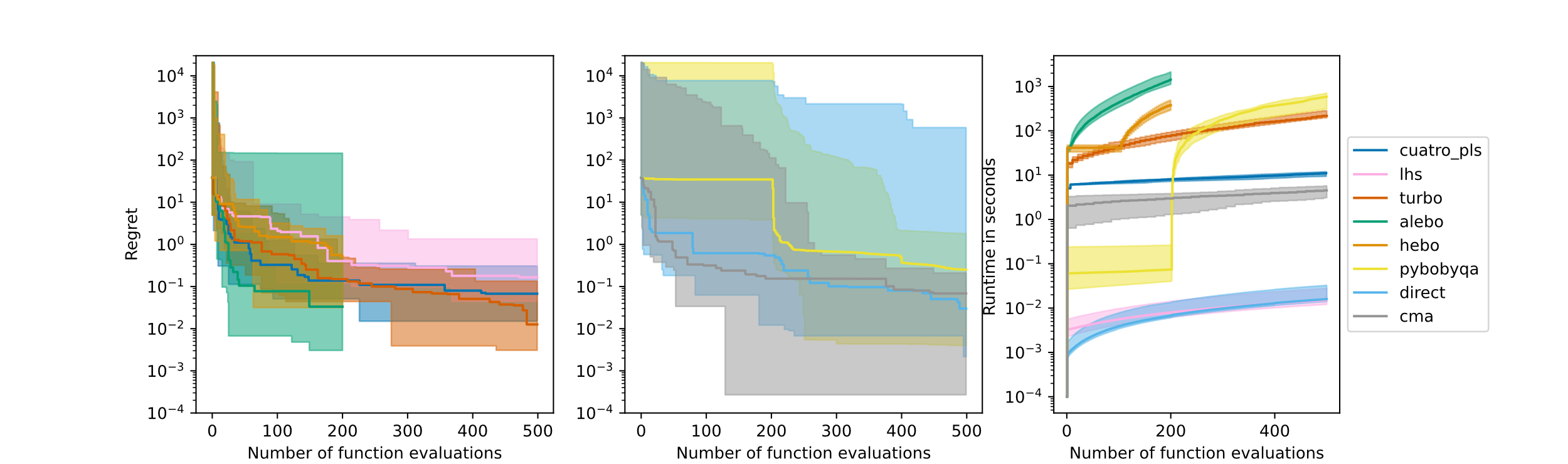
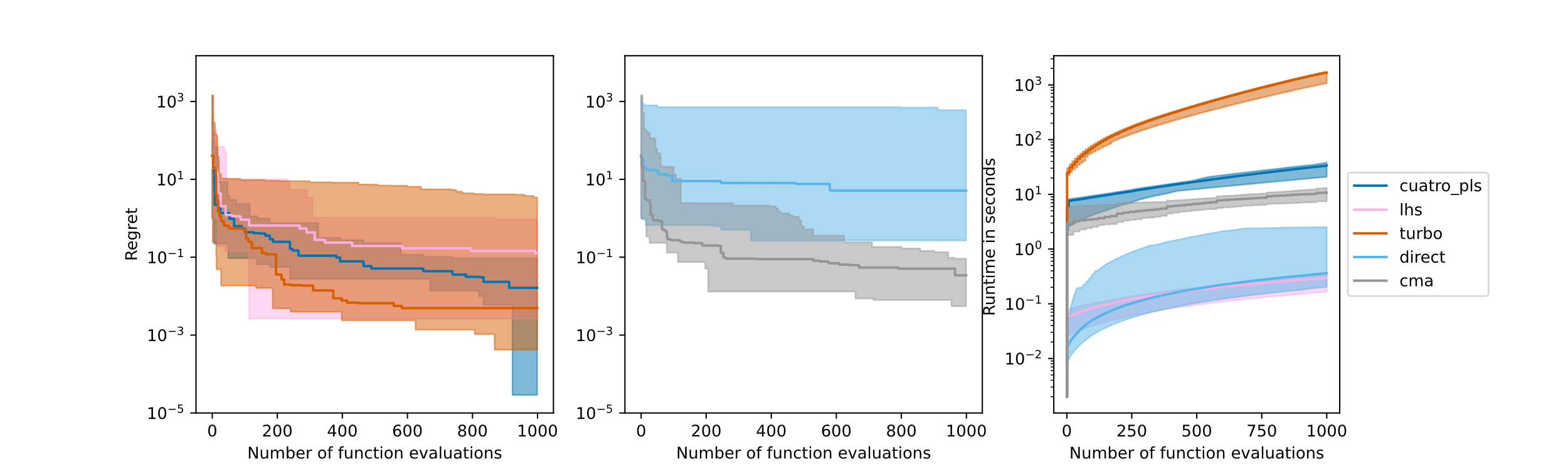
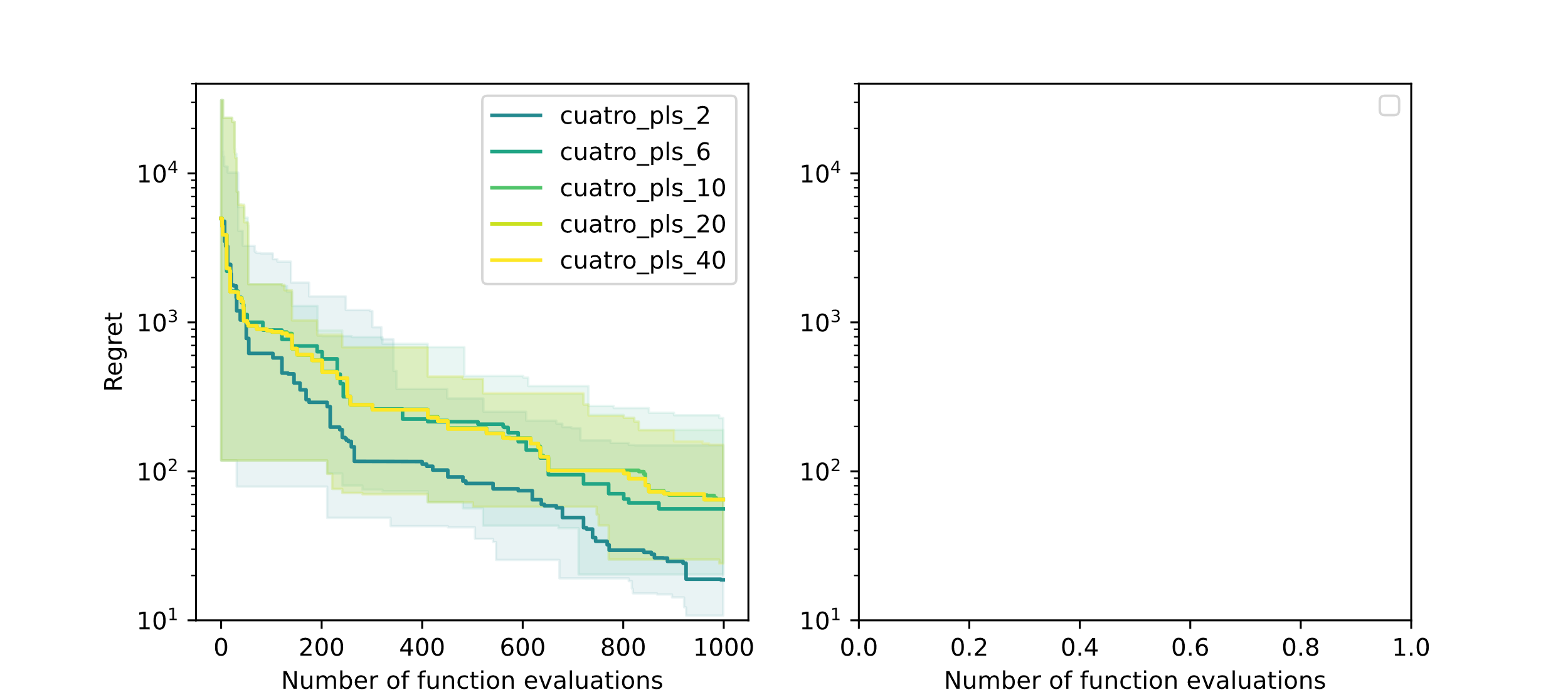


Figure 1. Regret versus evaluation budget convergence plot for the synthetic high-dimensional Rosenbrock case study with respective evaluation budget, original dimensionality and effective dimensionality of 500, 100, and 2 on the top and 1000, 1000 and 2 on the bottom.

Figure 2 investigates the effect that the input effective dimensionality has on CUATRO\_PLS performance with other hyperparameters fixed to default heuristics. We keep the RB original dimensionality at 1,000, and the *true* effective dimensionality at 10. We see no major difference in performance when perfectly or overestimating the true latent dimensionality at 6 or higher. To the contrary, the best performance is found when *underestimating* the effective dimensionality at 2 and 6. This suggests that CUATRO\_PLS stands as a powerful high-dimensional DFO solver, even if no knowledge about the intrinsic dimensionality is suspected. It might even be encouraged to use CUATRO\_PLS in fewer effective dimensions. In other words, there is merit to the idea of performing surrogate optimization in the *most* informative linear combination of dimensions for surrogate optimization, rather than capturing *all* information.

*Figure 2.* Regret versus evaluation budget convergence plot on the 1000-dimensional Rosenbrock with an effective dimensionality of 10 when CUATRO\_PLS is used with different input dimensionalities and all other hyperparameters fixed to default heuristics.

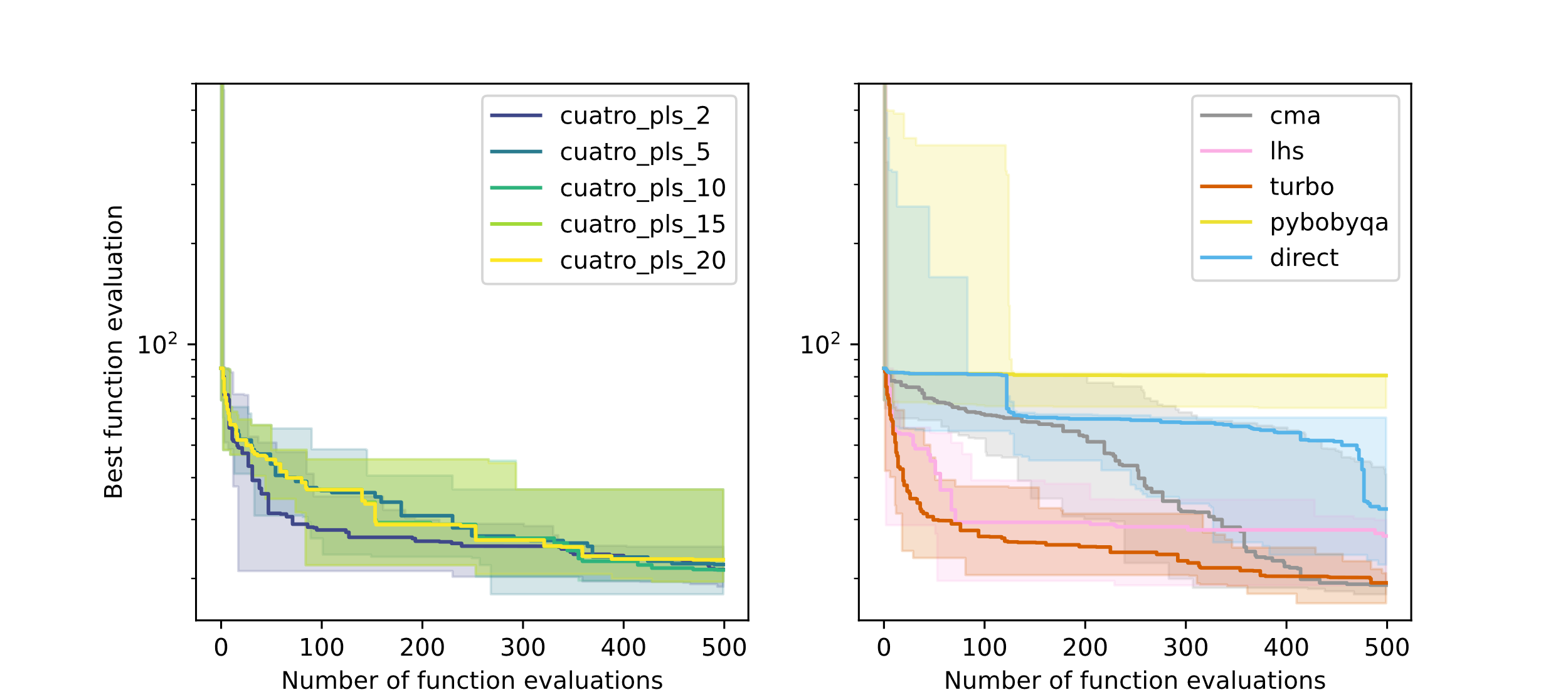
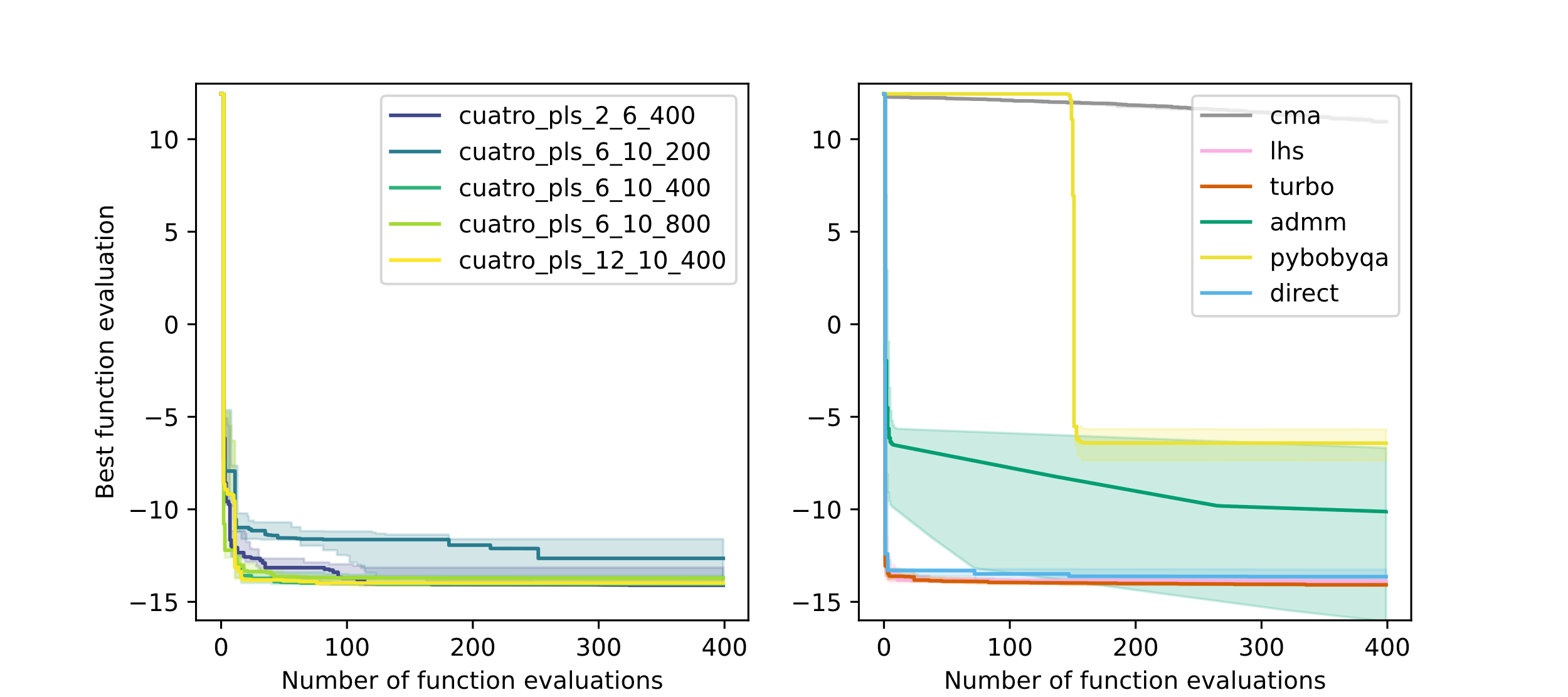
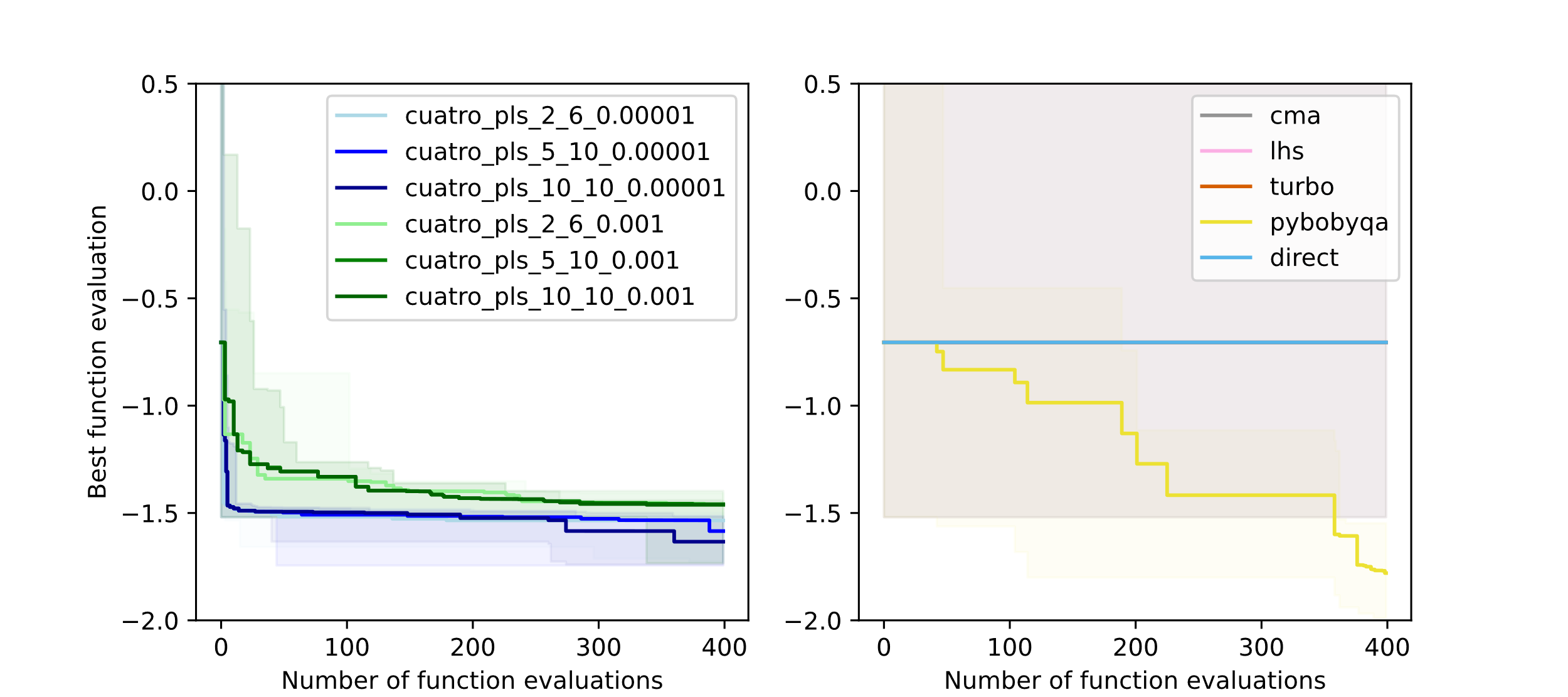
* + 1. Chemical engineering applications

Throughout all case studies in Figure 3, the CUATRO\_PLS runs (on the left) generally display consistency across different hyperparameters and input embedding dimensionalities. This supports our hypothesis that CUATRO\_PLS performs well even in the absence of problem-specific knowledge about embedding structure or effective dimensionality. In the policy training on the top, CUATRO\_PLS outperforms most methods and is only outperformed in consistency and in the median run by CMA-ES and TURBO respectively. In the value chain coordination in the middle, all CUATRO\_PLS median runs outperform the best runs of all other DFO methods, and even the original distributed optimization benchmark ADMM. In the hierarchical integration case study on the bottom, CUATRO\_PLS and Py-BOBYQA are the only methods that make any progress. While CUATRO\_PLS makes significant initial progress, Py-BOBYQA manages to better fine-tune the solution. Overall, the above findings are not surprising: CUATRO\_PLS, as a high-dimensional quadratic trust region optimizer, excels in finding *good* nearby optima *quickly*. However, CUATRO\_PLS remains a competitive optimizer even in the policy tuning, where more explorative methods like CMA-ES are encouraged.

* 1. Conclusion

We introduce CUATRO\_PLS as a scalable high-dimensional DFO solver that capitalizes on learned linear structures in the decision variables to display competitive performance with SOTA alternatives at significantly lower runtime compared to other model-based methods. Our solver converges on three high-dimensional chemical engineering DFO problems where intrinsic dimensionality is suspected, and excels in fine-tuning solutions, and finding good optima quickly. CUATRO\_PLS, by performing surrogate optimization in the most informative linear embedding dimensions, can even be used if no intrinsic dimensionality is known or suspected. In future work, we will combine our method with exploration routines to trade-off exploitation and exploration in the solution space.

Figure 3. Best function evaluation versus evaluation budget convergence plots on the policy training (top), value chain coordination (center), and hierarchical planning-scheduling (bottom) case studies. Left: CUATRO with different hyperparameters (embedding dimensionality, number of TR samples, and TR initial radius). Right: competitive solvers.



References

C. Cartis, L. Roberts, 2023, "Scalable subspace methods for derivative-free nonlinear least-squares optimization," Mathematical Programming, 199, 461–524.

R. Garnett, M. A. Osborne, P. Hennig, 2014, "Active learning of linear embeddings for Gaussian processes," In Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence (UAI), 230–239.

B. Letham, R. Calandra, A. Rai, E. Bakshy, 2020, "Re-Examining Linear Embeddings for High-Dimensional Bayesian Optimization," ArXiv, abs/2001.11659.

R. Moriconi, K. S. S. Kumar, M. P. Deisenroth, 2019, "High-dimensional Bayesian optimization with manifold Gaussian processes," arXiv preprint arXiv:1902.10675.

D. van de Berg, T. Savage, P. Petsagkourakis, D. Zhang, N. Shah, E. A. del Rio-Chanona, 2022,"Data-driven optimization for process systems engineering applications," Chemical Engineering Science, 248, 117135.

D. van de Berg, P. Petsagkourakis, N. Shah, E. A. del Rio-Chanona, 2023a, "Data-driven coordination of subproblems in enterprise-wide optimization under organizational considerations," AIChE Journal, 69(4), e17977.

D. van de Berg, N. Shah, E. A. del Rio-Chanona, 2023b, "Hierarchical planning-scheduling-control – Optimality surrogates and derivative-free optimization," arXiv:2310.07870

Z. Wang, F. Hutter, M. Zoghi, D. Matheson, N. De Feitas, 2016, "Bayesian optimization in a billion dimensions via random embeddings," Journal of Artificial Intelligence Research, 55, 361–387.